Pedagogical Self-Efficacy Improves Student Evaluations: Defining New Sub-Categories Across Faculty Gender in Management Education

Ted Ladd PhD*, Hult International Business School (San Francisco Campus),
Deepali D'Mello, Hult International Business School (Ashridge UK Campus)
Katarzyna Bachnik PhD, Hult International Business School (Boston Campus)
Amanda Nimon-Peters PhD, Hult International Business School (Dubai Campus)
Johan Roos PhD, Hult International Business School (Chelsea UK Campus)
Adam Reid, Hult International Business School (San Francisco Campus)

ABSTRACT

Pedagogical self-efficacy is the self-confidence that faculty members hold for their own ability to teach. This project used mixed method data from 20 interviews with faculty members, 85 self-assessments from professors, and 20,000 course evaluations from students to conclude that male professors show higher scores in three aspects of self-efficacy – designing a course, managing a classroom full of students, and providing feedback to students – than female professors. The data also show higher scores for male professors in student evaluations. Despite the gender differences, faculty self-efficacy for these three activities each predicted student evaluations for both male and female faculty members, with one exception. Male professors who reported higher confidence in their ability to design a course received lower evaluations from students. The distinctions have clear implications for both male and female professors, as well as for academic institutions seeking to develop their faculty and for researchers exploring pedagogical psychology.

Keywords: pedagogical self-efficacy, student evaluations, gender

* correspondence to ted@tedladd.com

INTRODUCTION

Student evaluations drive many aspects of academic institutions, such as course and program design, educational technology, campus architecture, institution rankings, student admissions, faculty hiring, and faculty promotion. Extant literature describes a significant role played by pedagogical self-efficacy (Bandura, 1997) in several of these aspects. Researchers have also repeatedly established the moderating effect of the teacher's gender on both pedagogical self-efficacy (Zhao, Seibert & Hills, 2005) and student evaluations (Rivera & Tilcsik, 2019).

This paper makes two contributions to this discussion. First, we identify and validate three granular attributes of pedagogical self-efficacy: ability to design a course; ability to manage students in the classroom; and ability to provide effective feedback to students. Our analysis reveals that these dimensions of pedagogical self-efficacy are all positive, significant predictors of student satisfaction as reported in student evaluations.

Our second contribution is a deeper examination of the moderating influence of faculty gender on the relationship between pedagogical self-efficacy and student evaluations. We find that male faculty members score higher on average than female faculty members for all three aspects of self-efficacy, as well as for student evaluations. Both genders show a positive, significant link between all three types of self-efficacy and student evaluations, with one exception. For male faculty members, higher scores for self-efficacy for course design predict a *decline* in student evaluations, perhaps because their over-confidence causes them to mis-judge student learning needs.

These results not only provide a foundation for subsequent research on pedagogical self-efficacy; they also illuminate ways in which academic institutions can bolster professors –

especially female members of the faculty – by increasing specific aspects of pedagogical self-efficacy to improve student satisfaction.

The paper is organized as follows. First, it reviews prior literature on self-efficacy, student evaluations, and the influence of gender across both variables. This review generates several testable hypotheses. Second, the paper describes the mixed methods employed in a two-phase data collection. In Phase 1, the theme of self-efficacy emerged during 20 interviews with faculty members from a large international business school. Phase 2 solicited self-assessments via survey from 85 professors. These data were mapped against the course evaluations from 20,000 students who had attended the surveyed professors' courses over the past two years. Third, the paper analyses the empirical findings from Phase 2 and discusses their implications. Finally, the paper concludes with commentary on the impact of these findings on the research and practice of pedagogy.

LITERATURE REVIEW AND HYPOTHESES

Pedagogical Self-Efficacy

Self-efficacy is a person's self-perception of her own ability to conduct a particular task (Bandura, 1977). Unlike the vernacular word "confidence," self-efficacy is domain specific. The construct predicts how often an agent will employ a specific behaviour, and even predicts how well the agent will use that behaviour to generate a result.

Several disciplines within management research have adapted and applied this theory (Staples, Hulland & Higgins, 1999 and Pillai & Williams, 2004). De Noble, Jung, and Ehrlich (1999) theorized and validated a construct of entrepreneurial self-efficacy (ESE) to reflect the entrepreneur's own assessment of the potential impact of their skills to start and operate a new venture. A person who believes that they can overcome challenges to complete a task is more likely to initiate and successfully complete that task. The ESE construct has been empirically linked to new venture performance in many circumstances (Chen, Greene, & Crick, 1998;

Hmieleski & Baron, 2008; McGee & Peterson, 2017). Responding to a call to explore the multi-dimensionality of ESE, McGee, Peterson, Mueller, and Sequeira (2009) proposed and validated four independent sub-dimensions that evaluate an entrepreneur's own perception of competence in common activities that entrepreneurs pursue as they launch a new venture: searching for a new idea; planning the design and launch of the idea; marshalling resources to pursue the idea; and implementing the plan in order to generate sustainable revenues. This research highlights the utility of applying the self-efficacy construct to other disciplines like pedagogy, and of defining some of the domain-specific activities within a discipline in order to explore more granular, actionable dimensions of self-efficacy.

The concept of pedagogical self-efficacy is over 30 years old (Guskey, 1982, 1988; Ashton. 1982; Tschannen-Moran & Hoy, 2001; Dellinger, Bobbett, Olivier & Ellett, 2008). Bandura himself (1997, 2001) developed a scale for teacher self-efficacy that included various tasks and attitudes. Gibson and Dembo (1984) proposed a construct and assessment tool to measure teacher self-efficacy. This was adapted for science teachers (Riggs & Enochs 1990; Bleicher, 2004) with the goal of providing academic institutions with an indicator for teacher effectiveness, allowing for earlier intervention and training if needed. A more recent study focusing on online classes (Corry & Stella, 2018) examined the sub-dimensions of professorial self-efficacy for instructional design, for managing the online classroom, and for facility with educational technology. Other studies (see Tschannen-Moran & Hoy, 2001; Dellinger et al., 2008) have also discussed measures for greater teacher self-efficacy.

Pedagogical self-efficacy emerged as a important theme our study. We began with qualitative interviews with faculty members at a large international business school to identify the key attributes of effective teaching. They not only focused on self-efficacy, but narrowed the concept to three specific dimensions of this concept: professorial self-confidence in a) designing an important course; b) managing a classroom filled with a large number of

internationally diverse students; and c) providing sufficient feedback to a large number of students on assignments they completed during the course.

This paper represents the culmination of the second phase of this project, which adds quantitative analysis to our qualitative data. The first hypothesis of this quantitative work seeks to determine the validity and reliability of constructs that were created to statistically reflect these three aspects of pedagogical self-efficacy.

Hypothesis 1: The dimensions of pedagogical self-efficacy for a) course design, b) classroom management, and c) providing feedback to students are valid and reliable scales that describe elements of pedagogical self-efficacy.

Student Evaluations

Student evaluations are ubiquitous in academia. The most common are those conducted at the conclusion of each course, where each student rates different aspects of the experience, from workload to professor passion to grading. Several studies have concluded that these evaluations are valid and reliable measures of the effectiveness of a professor's ability to teach content and skills to students (Centra, 1977, 1993; Cohen, 1981; Feldman, 1989; Koon & Murray, 1995; Marsh, 2007; Stehle, Spinath, & Kadmon 2012; Spooren, Brockx & Mortelmans, 2013). Marsh (1987) contended that student evaluations were the only indicator of teaching effectiveness whose validity has been thoroughly and rigorously established. Extant literature addresses several classroom techniques (Apperson, Laws & Scepansky, 2006; Jenkins, 2013; Touchton, 2015) and professional behaviours (Spooren & Mortelmans, 2006; Hackman & Walker, 2009; Evans & Waring, 2011) that professors can adopt to increase the scores of their student evaluations and thereby improve student learning outcomes.

There is controversy, however, about the assertion that student evaluations reflect student learning. Some researchers worry that evaluations suffer from recency bias (Steiner & Rain, 1989; Dickey & Pearson, 2005) and a personal affection for certain professors (Ambday &

Rosenthal, 1993; Hamermesh & Parker, 2005). Evaluations might also reflect a professor's low expectations for student preparation, easy assignments (Centra, 2003), and high grades (Gump, 2007; McPherson, 2006), all of which are presumably favoured by students. This paper does not attempt to address or resolve these controversies, but there is little doubt that, regardless of the effectiveness of student evaluations to capture the long-term impact of a course on their learning and career outcomes, academic institutions, accreditors, and ranking agencies take them very seriously. In short, improving student evaluations is an important task for any professor, regardless of the link between evaluations and learning.

Just as self-efficacy is a reliable predictor of performance in other management disciplines, we posit that each of the three aspects of pedagogical self-efficacy predicts student evaluations.

Hypothesis 2: Pedagogical self-efficacy for a) course design, b) classroom management, and c) providing feedback to students are significantly and positively related to student evaluations of the course.

The Moderating Influence of Gender

Student evaluations show a well-known bias towards male professors and against female professors (Basow, 1995; Centra & Gaubatz, 2000; Young, Rush, & Shaw, 2009; Boring & Stark, 2016; Rivera & Tilcsik, 2019). Self-efficacy shows a similarly skewed distribution for gender. (See Ladd, Hind, and Lawrence (2018) for a recent example and Zhao, Seibert, and Hills (2005) for a summary and meta-analysis.) Several studies have concluded that the mechanism of self-efficacy operates differently for men than for women. For example, BarNir, Watson, and Hutchins (2011) concluded that women are more influenced by role models than men.

In the final step in our analysis, we posit as a third hypothesis that the relationships between pedagogical self-efficacy and student evaluations in Hypothesis 2 are positive and significant for both men and women.

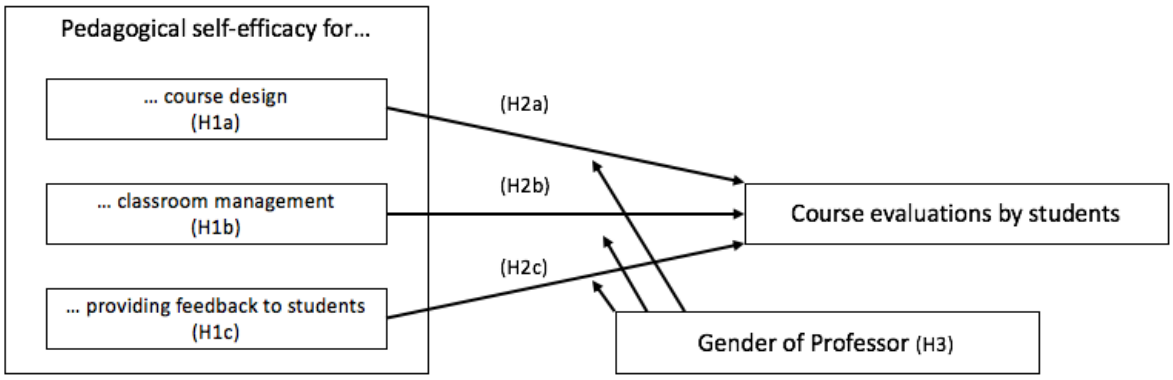
Hypothesis 3: The relationship between the three aspects of pedagogical self-efficacy – a) course design, b) classroom management, and c) providing feedback to students – and student evaluations are significantly moderated by faculty gender.

These hypotheses are summarized in Table 1 and depicted in Figure 1.

Table 1: Summary of Hypothesis

Variable	**Hypothesis 1**	**Hypothesis 2**	**Hypothesis 3**
a) Pedagogical self-efficacy for course design…	…is a valid construct and…	…predicts student evaluations…	…when moderated by professor gender.
b) Pedagogical self-efficacy for classroom management…	…is a valid construct and…	…predicts student evaluations…	…when moderated by professor gender.
c) Pedagogical self-efficacy for providing feedback…	…is a valid construct and…	…predicts student evaluations…	…when moderated by professor gender.

Figure 1: Depiction of Hypotheses



METHODOLOGY

This mixed method study was conducted in two phases: a qualitative exploration of the drivers of student evaluation as reported by professors, and then a statistical analysis that develops the empirical relationships between pedagogical self-efficacy, student evaluations, and gender.

Phase 1 Procedure and Participants

We interviewed 20 professors – 15 men and 5 women – in 2018 at a large international business school. These faculty lived and taught in five different campuses across four different countries in three different continents. These faculty taught business courses at the postgraduate level (e.g. Master of Business Administration). Our interview pool consisted of a cross-section of teachers across all of the school's international campuses and included those who had won teaching awards as an exercise in purposive opportunity sampling (Collingridge & Gantt, 2008). We did not include teachers in the sample who had been recently hired or recently fired. In this paper, we define a professor as the person who delivers a course. We are not restricting this definition to those who conduct research or have secured tenure.

We used a semi-structured interview technique in which our goal was to understand the professor's unique practices before, during and after a class session to determine what these faculty members believed to be the drivers of student satisfaction and thereby student evaluation of each course. The interviews typically lasted from 30 to 60 minutes and were conducted on audio or video Zoom by four members of the research team. All interviews were conducted in English, recorded and then transcribed verbatim. Informed consent was taken and right to withdraw was explained.

We used thematic analysis to analyse the data using the structural coding approach outlined by Saldana (2011). Using Dedoose software (v 4.12), we applied codes to excerpts within the interview transcripts. The coding was conducted in two stages in a bottom-up, inductive approach (Charmaz, 2014). The first stage involved reflecting line by line on paragraphs in the interview transcripts to interpret the underlying meaning (decoding). Paragraphs were then labelled with a code, which represented the meaning we identified (encoding). In this way, 'concepts' of repeatedly expressed ideas in professors' practices before and during the classroom were produced and then clustered into categories.

The iterative process of classifying concepts and organizing them by categories continued until the data yielded no further codes and categories. The interviewers did not conduct the initial coding but did review the codes alongside the original transcripts (Charmaz, 2014; Urquhart, 2001). The final coding and categorising were concluded following discussion and agreement between the interviewers and coders (Syed & Nelson, 2015). The thematic analysis incorporated not only 'what' professors were doing before and during the classroom sessions, but 'why' they are doing that (Urquhart, Lehmann, & Myers, 2010).

The main themes from these interviews revolved around designing courses, communicating course deliverables, managing the classroom, optimizing the methods for sharing content, soliciting student discussion in classroom, enforcing punctuality, providing feedback to students, and participating in student experiences outside of class. The concept of self-efficacy emerged organically from these interviews; the authors did not intentionally probe for these opinions. Several respondents declared that self-confidence supported the three key dimensions i.e. course design, classroom management, and providing feedback to students.

Phase 2 Procedure and Participants

This phase involved the creation and dissemination of a survey to all professors in the postgraduate programs of the same international business school about their attitudes, beliefs, behaviours, and techniques with regards to teaching. Many of the questions in the survey emanated directly from the Phase 1 qualitative study.

The survey was sent out via email to 152 professors across all global campuses. Ninety-two participants completed it. Because course evaluation data for six of the participants were not available, and two participants only partially completed the survey, we used data of N=84 participants for the full analysis yielding a response rate of 55%. Of the 84 respondents, 16 participants self-identified as female (19%) and 68 as male (81%). (Our survey offered respondents an opportunity to identify with other gender identities. None did; all respondents

self-identified as either male or female.) Forty-seven participants (56%) were adjuncts and 37 were full-time (44%) faculty. We used IBM SPSS 23, AMOS 25 and Hayes Process Macro Plugin Version 3.4 to conduct the exploratory factor analysis (EFA) for the self-efficacy constructs.

Separately, we collated the responses from these professors' student evaluations across all campuses of the business school. For the purpose of this study, only postgraduate student evaluations of professors for the academic year 2017-18 and 2018-19 were used in order to minimise memory recall issues on part of the professors. A total of N=20,339 student evaluations were gathered from 1,362 students who had been in a class taught by one of these 84 professors. 55% percent of the students were male and 45% were female. (The student demographic collection method at the time of their intake into the program did not consider other gender identities.) Their average age at the time of the survey was 27 years. The school's student evaluation form contained 15 questions for the academic year 2017-18 and 18 questions for academic year 2018-19. Students responded to these questions on a 5-point scale Likert scale ranging from 1 being strongly disagree to 5 being strongly agree. This study only used the question that asked students for an evaluation of the course.

The student surveys were submitted anonymously. While this improved the validity of the evaluations, it also eliminated any opportunity to explore the demographic characteristics of the students to note any differentiations.

RESULTS

Table 2 presents the descriptive statistics along with the correlations between the variables under study. Course evaluations and the three aspects of self-efficacy were reported on a 5-point Likert scale. Faculty gender was recorded as 1 for female and 2 for male.

Table 2: Mean (M), Standard Deviation (SD) and Correlations of the study variables, N=20339

Variable	*M*	*SD*	1	2	3	4
1. Course Evaluation	4.19	1.03				
2. Self-efficacy for Course Design	4.83	.370	-.02**			
3. Self-efficacy for Classroom Mgmt.	4.54	.550	.04**	.55**		
4. Self-efficacy for Providing Feedback	4.63	.580	.08**	.42**	.47**	
5. Faculty Gender	1.84	.370	.03**	.04**	.00	.08**

** p<0.10

Pedagogical Self-Efficacy

Table 3 presents the results of the exploratory factor analysis for the three aspects of pedagogical self-efficacy relating to a) course design, b) classroom management and c) providing feedback. The number of factors were not set before the EFA was conducted. It was determined a priori that any item with a factor loading greater than .40 would be accepted, as suggested by Field (2018). We used principal component analysis as the method for extraction and all 35 items were subjected to an EFA with varimax rotation. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .80 (Hair, Black, Babin, & Anderson, 2013). Bartlett's test of sphericity = 944.733, $p < .001$ indicated that correlation structure is adequate for factor analyses (Hair et al., 2013). Our maximum likelihood factor analysis employed a minimum threshold correlation of .40 and Kaiser's criterion of eigen values greater than 1 (Stevens, 1992; Field, 2018) to generate the best fit for the data, accounting for 39.04% of the variance for self-efficacy for course design, 37.23% of the variance for self-efficacy for classroom management and 55.87% of the variance for self-efficacy for feedback. A total of 25 items (Table 3) across the three dimensions were selected as a result of high factor loading for their respective factors: 4 items were retained for course design, 15 for managing the classroom and 6 for providing feedback.

We established face validity through both the Phase I portion of the research project and a consensus among the study's multiple authors about which questions to insert into the faculty-oriented survey.

We confirmed reliability using Cronbach's alpha (Hair et al., 2013). For course design the reliability was α= .65, for classroom management it was α=.87 and for providing feedback it was α=.78. These are acceptable reliability values and indicate moderately high reliability for each of the constructs (Cronbach & Meehl, 1955). These results support Hypotheses 1a, 1b and 1c.

Table 3: Items for Pedagogical Self-Efficacy

I am confident in my ability to…

Item	Factor 1: Self-Efficacy for Course Design	Factor 2: Self-Efficacy for Classroom Management	Factor 3: Self-Efficacy for Feedback
...help students achieve the learning objectives.	.78		
...keep students engaged in the topic.	.70		
...help students learn relevant skills for global managers.	.42		
...set high standards for student achievement.	.73		
...control the intellectual discussion in the room.		.71	
...manage technology effectively.		.54	
...inspire and motivate students to learn the material		.56	
...create a sense of community and belonging		.65	
...help students achieve the learning objectives for the day		.55	
...set high standards for student achievement		.56	
...address a large audience		.62	
...accommodate the range of individual differences among students		.58	
...successfully maintain a positive classroom climate		.52	
...navigate different student learning styles		.73	
...build a sequence of activities to achieve the learning objective		.62	
...incorporate students' prior knowledge		.57	
...overcome disruptive student behavior		.70	
...provide significant challenges for my most capable students		.50	
...give an alternate explanation or example when students are confused		.60	

I am confident in my ability to…

Item	Factor 1: Self-Efficacy for Course Design	Factor 2: Self-Efficacy for Classroom Management	Factor 3: Self-Efficacy for Feedback
…offer insightful feedback			.80
…offer timely feedback			.72
…offer clear feedback			.79
…offer constructive feedback			.84
…offer motivating feedback			.77
…offer blunt feedback			.49
Eigen value	1.952	5.585	3.352
% of Total Variance	39.046	37.233	55.870
Cronbach's Alpha	.65	.87	.78

Note: Extraction Method: Principal Axis Factoring**;** Rotation Method: Varimax with Kaiser Normalization

Pedagogical Self-Efficacy and Student Evaluations

Simple regression was carried out to test hypothesis 2. The three different regression models are presented in Table 4, Table 5 and Table 6 to explore the relationship between each aspect of pedagogical self-efficacy and student evaluation. The column in the middle of each table ("Term p-value") reports significance of each term in each model. The column in the far right of the table ("Model p-value") reports the significance of each model, which includes the constant and the independent variable.

Despite small effect sizes, all of the term *and* model p-values are significant ($p<.001$), driven by the large sample size of the study. In this and all subsequent analyses, the coefficients are unstandardized, and we used 95% to calculate the lower level (LLCI) and upper level (ULCI) for the confidence intervals.

Table 4: Regression Analyses for Pedagogical Self-Efficacy for Course Design Predicting Student Evaluations

Variable	Coeff	S.E.	t-value	Term p-value	LLCI	ULCI	R^2	F	Model p-value
Constant	4.54	0.09	48.61	0.000	4.35	4.72			
Self-Efficacy for Course Design	-0.07	0.02	-3.74	0.000	-0.11	-0.03	0.001	14.02	0.000

Table 5: Regression Analysis for Pedagogical Self-Efficacy for Managing the Classroom Predicting Student Evaluations

Variable	Coeff	S.E.	t-value	Term p-value	LLCI	ULCI	R^2	F	Model p-value
Constant	3.87	0.06	64.77	0.000	3.75	3.99			
Self-Efficacy for Manage Classroom	0.07	0.01	5.43	0.000	0.05	0.10	0.001	29.49	0.000

Table 6: Regression Analyses for Pedagogical Self-Efficacy for Providing Feedback Predicting Student Evaluations

Variable	Coeff	S.E.	t-value	Term p-value	LLCI	ULCI	R^2	F	Model p-value
Constant	3.51	0.06	61.05	0.000	3.40	3.63			
Self-Efficacy for Providing Feedback	0.15	0.01	11.86	0.000	0.12	0.17	0.007	140.70	0.000

The results of the Pearson's correlation reported in Table 2 align with the results from simple regression in Table 4, Table 5 and Table 6 to conclude that the relationship between self-efficacy for course design is significant and negatively related to student evaluations, requiring that we reject Hypothesis 2a. This indicates that participants who rated themselves higher on course design had lower course evaluations. Please note, this does not mean having high self-efficacy in course design causes lower student evaluations. In contrast, self-efficacy for classroom management shows a significant and positive relationship with student evaluations, supporting Hypothesis 2b. Similarly, self-efficacy for providing feedback shows a significant and positive relationship with student evaluations, supporting Hypothesis 2c.Therefore

participants who rated themselves to have higher self-efficacy for classroom management and providing feedback had higher course evaluations. It is important to note that while all relationships were significant, the effect sizes and the explained variances were small suggesting other variables also play important roles in determining student evaluations.

The Moderating Influences of Gender

Table 7 reports the descriptive statistics of the three different aspects of pedagogical self-efficacy, as well as student evaluations, by gender. This table also reports the ANOVA statistics to compare the means across gender. Self-efficacy for course design and for providing feedback show significant differences across gender, where male faculty members show higher scores than female faculty members. Self-efficacy for classroom management does not show a significant difference across gender. Student evaluations of the course also show a significant difference, where male faculty members earn almost 0.1 more than female faculty members on average for student evaluations out of 5.0. (To put this latter statistic is starker terms, 83% of students rated a course either a 4 or a 5; only 17% of students rated a course 1, 2, or 3. Thus the effective range for student evaluations is really from 4.0 to 5.0. The difference of 0.1 over an effective range of 1.0 means that male professors on average earn a 10% higher course evaluation than female professors.)

Table 7: Descriptive Statistics and ANOVA Across Professor Gender

	Male Faculty			Female Faculty			ANOVA	
Type of Self-Efficacy:	**Mean**	**S.D.**	**N**	**Mean**	**S.D.**	**N**	**F**	**p-value**
Course Design	4.84	0.37	16990	4.80	0.40	3349	25.5	0.000
Classroom Management	4.54	0.55	16888	4.55	0.63	3074	0.2	0.639
Providing Feedback	4.65	0.58	16888	4.52	0.63	3074	129.0	0.000
Student Evaluations	4.20	1.04	16888	4.11	1.00	3349	23.1	0.000

As noted, Table 4 above explored the relationship between pedagogical self-efficacy and course evaluation for our entire sample. Our third hypothesis inquires as to whether the relationships between each aspect of pedagogical self-efficacy and student evaluations are the

same for each gender. The ANOVA results in Table 7 suggest that gender might confound those relationships. To explore this, we conducted three separate moderated regression analyses.

Self-Efficacy for Course Design and Student Evaluations Moderated by Gender

Table 8 reports the results of the moderation analysis for pedagogical self-efficacy for course design and student evaluation. The model without the interaction term explains only 1.2% of the variation (R squared), but this is nonetheless significant ($p<.001$; $F=83.2$; df 3, 20301). The addition of the interaction between self-efficacy and gender increases the R squared to 2.24% ($p<.001$; $F=201.8$). Course evaluations and self-efficacy were reported on a 5-point Likert scale. Faculty gender was recorded as 1 for female and 2 for male.

Table 8: Self-Efficacy for Course Design and Student Evaluations Moderated by Gender

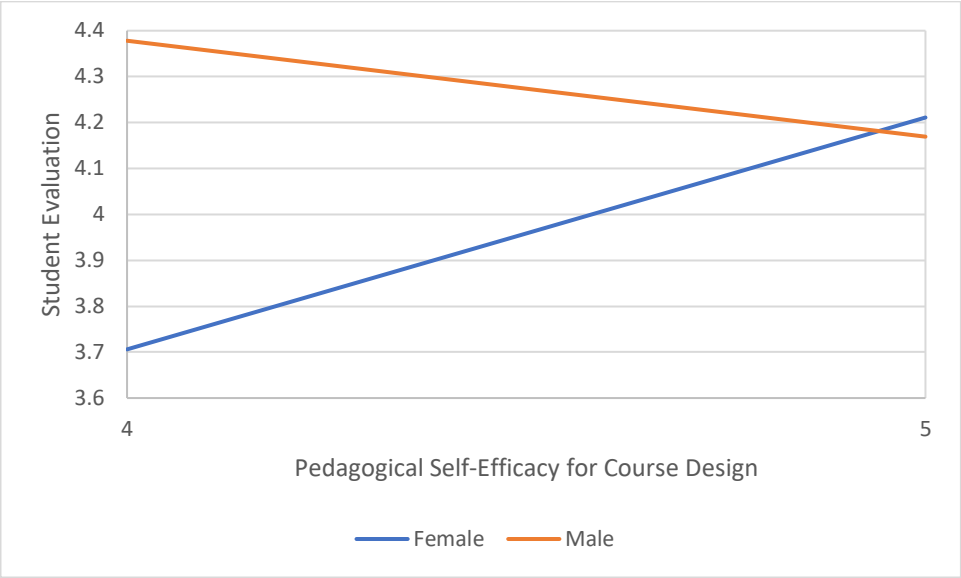
	Coeff	**S.E.**	**t-value**	**p-value**	**LLCI**	**ULCI**
Constant	-1.84	0.44	-4.19	0.000	-2.70	-0.98
Self-efficacy for course design	1.22	0.09	13.37	0.000	1.04	1.40
Gender	3.53	0.24	14.88	0.000	3.06	3.99
Interaction	-0.71	0.05	-14.52	0.000	-0.81	-0.62

The conditional effects in Table 9 describe the relationship between self-efficacy and student evaluations for male or female faculty members separately. Both effects are significant ($p<.001$), but different. For female faculty members, self-efficacy for course design directly predicts student evaluations, where a rise in one generates a rise in the other. For male faculty members, self-efficacy for course design is inversely related to student evaluations. These contrasts are illustrated in Figure 2. This conclusion causes us to reject Hypothesis 3a.

Table 9: Conditional Effects of Gender on Self-Efficacy for Course Design and Student Evaluations

Gender	**Effect**	**S.E.**	**t-value**	**p-value**	**LLCI**	**ULCI**
Female	0.50	0.04	11.39	0.000	0.42	0.59
Male	-0.21	0.02	-9.83	0.000	-0.25	-0.17

Figure 2: The Relationship between Self-Efficacy for Course Design and Student Evaluations Moderated by Gender



Self-Efficacy for Classroom Management and Student Evaluations Moderated by Gender

Table 10 reports the results of the moderation analysis for pedagogical self-efficacy for classroom management and student evaluation. The model without the interaction term explains only 0.44% of the variation (R squared), but this is nonetheless significant ($p<.001$; $F=29.27$; df 3, 19924). The addition of the interaction between self-efficacy and gender increases the R squared to 0.61% ($p<.001$; $F=34$).

Table 10: Self-Efficacy for Classroom Management and Student Evaluations Moderated by Gender

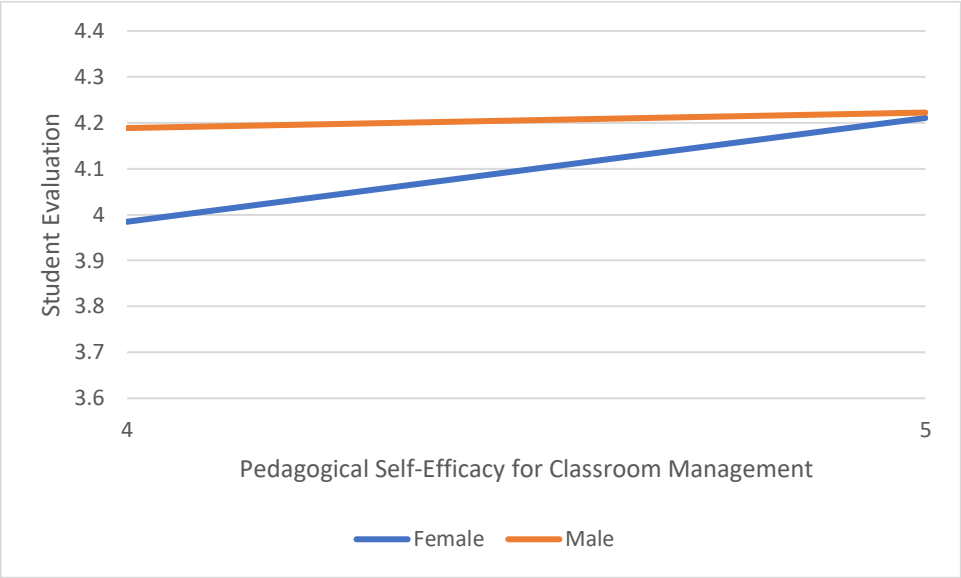
	Coeff	**S.E.**	**t-value**	**p-value**	**LLCI**	**ULCI**
Constant	2.11	0.28	7.54	0.000	1.56	2.66
Self-efficacy for course design	0.42	0.06	6.86	0.000	0.30	0.54
Gender	0.97	0.15	6.44	0.000	0.68	1.27
Interaction	-0.19	0.03	-5.83	0.000	-0.26	-0.13

The conditional effects in Table 11 conclude that both male and female faculty members show a significant and positive relationship between self-efficacy for classroom management and student evaluations, supporting Hypothesis 3b, illustrated in Figure 3.

Table 11: Conditional Effects of Gender on Self-Efficacy for Classroom Management and Student Evaluations

Gender	Effect	S.E.	t-value	p-value	LLCI	ULCI
Female	0.23	0.03	7.64	0.000	0.17	0.28
Male	0.03	0.01	2.33	0.020	0.01	0.06

Figure 3: The Relationship between Self-Efficacy for Classroom Management and Student Evaluations Moderated by Gender



Self-Efficacy for Providing Feedback and Student Evaluations Moderated by Gender

Table 12 reports the results of the moderation analysis for pedagogical self-efficacy for providing feedback to students and the students' evaluation of the course. The model without the interaction term explains only 0.81% of the variation (R squared), but this is nonetheless significant ($p<.001$; $F=54.46$; df 3, 19924). The addition of the interaction between self-efficacy and gender slightly increases the R squared to 0.84% ($p<.01$; $F=6.73$).

Table 12: Self-Efficacy for Providing Feedback and Student Evaluations Moderated by Gender

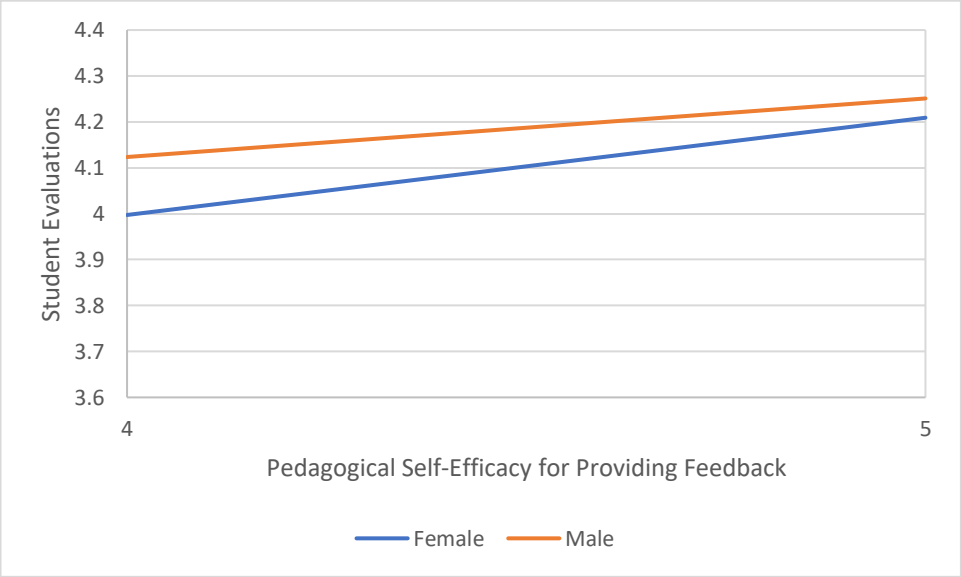
	Coeff	S.E.	t-value	p-value	LLCI	ULCI
Constant	2.69	0.28	9.72	0.000	2.15	3.23
Self-efficacy for course design	0.30	0.06	4.89	0.000	0.18	0.41
Gender	0.46	0.15	3.11	0.002	0.17	0.76
Interaction	-0.08	0.03	-2.60	0.010	-0.15	-0.02

The conditional effects in Table 13 conclude that both male and female faculty members show a significant and positive relationship between self-efficacy for providing feedback to students and the students' evaluations of the course, supporting Hypothesis 3c, illustrated in Figure 4.

Table 13: Conditional Effects of Gender on Self-Efficacy for Providing Feedback and Student Evaluations

Gender	Effect	S.E.	t-value	p-value	LLCI	ULCI
Female	0.21	0.03	7.18	0.000	0.15	0.27
Male	0.13	0.01	9.36	0.000	0.10	0.15

Figure 4: The Relationship between Self-Efficacy for Providing Feedback and Student Evaluations Moderated by Gender



Conclusion from Analyses

The statistical conclusions from these hypotheses are summarized in Table 14 and depicted in Figure 5. Hypotheses supported by our analyses are highlighted in green. Rejected hypotheses are highlighted in red.

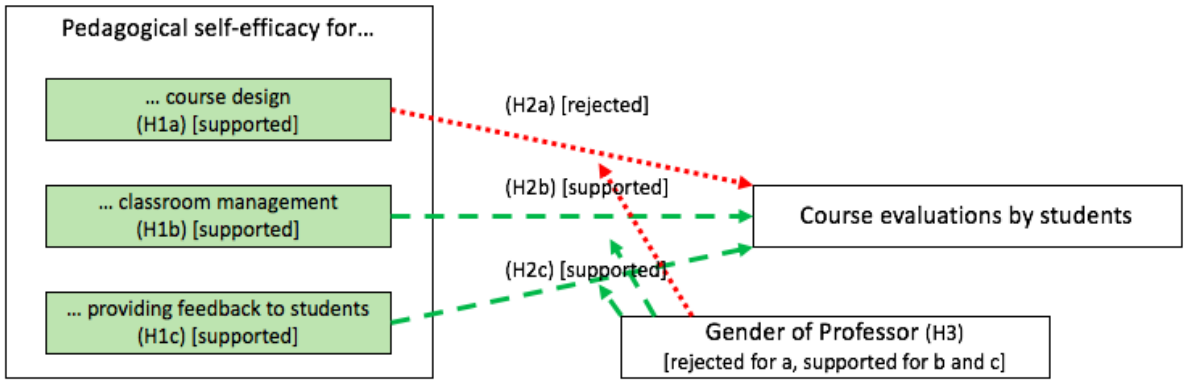
We can also see that almost all of our analyses show a significant, positive link between pedagogical self-efficacy and student evaluations. The one exception – where male professors

show a negative relationship between self-efficacy for course design and student evaluations – is driving the rejections of both Hypotheses 2a and 3a.

Table 14: Summary of Conclusions

Variable	Hypothesis 1	Hypothesis 2	Hypothesis 3
a) Pedagogical self-efficacy for course design…	…is a valid construct and… [supported]	…predicts student evaluations… [rejected]	…when moderated by professor gender. [rejected]
b) Pedagogical self-efficacy for classroom management…	…is a valid construct and… [supported]	…predicts student evaluations… [supported]	…when moderated by professor gender. [supported]
c) Pedagogical self-efficacy for providing feedback…	…is a valid construct and… [supported]	…predicts student evaluations… [supported]	…when moderated by professor gender. [supported]

Figure 5: Depiction of Conclusions



DISCUSSION

Pedagogical self-efficacy refers to the self-confidence that faculty members hold for their own ability to teach. This paper defined and explored three specific aspects of pedagogical self-efficacy regarding the faculty member's self-confidence: a) to design a course; b) to manage the classroom environment during a class session; and c) to provide feedback to students. We found that all three of our constructs are reliable and valid across a sample of 85 faculty members.

Most of these constructs show substantial differences across male and female faculty members, where men have higher scores for self-efficacy for course design and self-efficacy

for providing feedback. Male teachers also receive about 10% higher course evaluations from students than female teachers. Self-efficacy for classroom management does not show a difference across faculty gender.

Two of these constructs – self-efficacy for classroom management and self-efficacy for providing feedback – predict student evaluations, where a higher score for self-efficacy relates to higher student evaluations. These relationships are slight, with small effect sizes. However, they are all significant, driven by a sample size of about 20,000 student evaluations. These relationships hold for both male and female faculty members.

The third construct – self-efficacy for course design – shows a significant but negative relationship for male faculty with student evaluations, where a higher score for male pedagogical self-efficacy for course design predicts a lower evaluation. Female faculty members show a significant positive relationship between self-efficacy for course design and student evaluations.

CONCLUSION

The conclusions of this analysis have several implications for teachers, students, academic institutions, and researchers.

Implications for Faculty Members

Teachers who seek to improve their evaluations from students have many different approaches and techniques to utilize: setting moderately challenging but attainable goals to help students experience mastery first-hand; finding distant role models and proximate mentors in the specific target domain; practicing self-reflection; and embracing optimism even in the face of failure to become more resilient (Bandura, 2008).

Instead of focusing on what techniques the professor could employ to boost student evaluations, this paper explored the attitudes of professors, specifically focusing their self-confidence to effectively perform three key activities: designing a course, managing the

classroom, and providing feedback to students. We created and validated scales for each of these constructs of pedagogical self-efficacy and established a significant link between the professors' self-confidence in these domains with the students' evaluation of the professors' courses.

These analyses concluded that teachers should seek to build self-efficacy in each of three areas in order to improve student evaluations. For example, a professor might find support from three different mentors: one who has achieved success in designing a course, one who manages the classroom well, and a third whose mechanisms for providing feedback to students has been well-regarded. In addition to improving a faculty member's overall course experience, and thus student evaluations, these interactions will boost the faculty member's own self-confidence in these domains.

There is an exception to this conclusion. Male faculty members should exert care in bolstering their self-efficacy specifically for designing a new course. Higher self-confidence in this domain leads to *lower* student evaluations. If this inverse relationship had appeared in self-efficacy for classroom management or providing feedback, we might have assumed that students perceived male professors as over-confident. In fact a recent study by Martin & Phillips (2017) reported differences in workplace confidence, especially in male-dominated environments (e.g., business school) and positions of power (managerial positions). However, since course design occurs out of direct view of students, it is possible that this inverse relationship is caused by the male faculty member's selection of course objectives and materials that are misaligned with student expectations. If this is accurate, male faculty members might alleviate some of this problem by piloting and benchmarking course designs prior to full implementation.

Counter-intuitively, one possible way for female faculty to bolster their pedagogical self-efficacy is to change their interpretation of the benchmark for student evaluations. Here is the

logic of this paradoxical recommendation. Female professors consistently receive lower evaluations than male professors (Boring & Stark, 2016; Rivera & Tilcsik, 2019). The likelihood of that would stem from the possibility that students *perceive* female professors as less effective, even if the students are receiving the same quality of education and same learning outcomes as they receive from male professors. Much research has indicated that women's performance and qualifications are under-rated compared to their male counterparts under identical conditions when the gender of the person being rated is known (e.g. Isaac, Lee, & Carnes, 2009; Macnell, Driscoll, & Hunt, 2014; Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012; Terrell et al., 2017). Although this effect can disappear when gender is hidden, the gender identity of a professor is an unavoidable condition in teaching evaluations of classroom-based courses. Therefore, the difference in course evaluation ratings between those given to male and female professors is more likely driven by unconscious bias in student ratings than any genuine difference in ability. Female professors may be facing a headwind of student perception that unfairly lowers their evaluation.

Failure to recognize this bias might cause female faculty members to see their lower evaluations as a humbling commentary on their ability. Instead, acknowledging an inherent bias in student evaluations might help female faculty members better appreciate their own talents and contributions. Such a perspective might improve their own pedagogical self-efficacy, which in turn would actually positively influence the student evaluations. This acknowledgement would also alert male faculty members to the potential for bias that might inflate their own evaluations.

Implications for Academic Institutions

Some academic institutions provide training courses for their faculty in the hope of boosting student evaluations – and of course the students' learning outcomes that these evaluations hopefully reflect. This paper suggests that this approach to faculty training may not be the other

route to improve student experiences. Instead of or in addition to providing guidance to faculty on specific teaching techniques, academic institutions could bolster their faculty members' sense of self-efficacy for three distinct aspects of the classroom experience: course design, classroom management, and providing feedback to students. This will require a more personalized approach to faculty development, where the administrators adjust goals and mentoring possibilities to align with each faculty member's current sense of self-efficacy for these three activities. This individual attention also obviates any systematic differentiation by gender.

The link between self-efficacy and student evaluation has a second, more controversial impact on academic institutions: hiring new faculty members. An approach to recruiting faculty members that includes an assessment of a candidate's self-confidence in the three pedagogical activities would have two negative consequences. The first is based on this paper's empirical conclusions. Hiring male faculty members who show high scores for self-efficacy for course design would, on average, lower student evaluations.

The second consequence is a longer-term projection on gender diversity, as a focus on self-efficacy might elevate male candidates over female candidates. To the extent that self-efficacy derives from the availability of role models, using it as a measure for hiring could give preference for male professors over female professors. This would reduce the number of successful female teachers to act as role models for subsequent generations of female teachers, further demoralizing female self-efficacy. To compound this negative spiral, students would encounter fewer female faculty members, which might amplify their own gender bias in evaluating courses taught by female professors, widening the differences in student evaluations between male and female professors. Although self-efficacy is likely to continue to drive student evaluations, a long-term goal to eliminate the gender gap for both of these variables will help overcome short-term tactics that might exacerbate these differences.

Implications for Researchers

This paper has several implications for researchers in education. First, by separating pedagogical self-efficacy into three separate, distinct, valid, reliable constructs that form part of the sequence of delivering an educational experience to students, this paper provides a foundation for other researchers to further deconstruct the pedagogical path and assemble measures of self-efficacy for each of them. For example, this paper did not explore activities that occur outside of the classroom, such as teacher-student coaching. Nor did it explore self-efficacy for specific teaching methods, like the Flipped Classroom (where teachers pre-record their lecture for students to watch before class, and then spend the class session in discussion). Yet both of these activities might be significant contributors to student evaluations with their own, distinct measures of self-efficacy.

Second, this paper showed the links between the three aspects of pedagogical self-efficacy and student evaluation across both genders but did not explore the mechanisms by which the teachers' self-confidence generates better student experiences. For example, a comparison of the regressions in Table 4, Table 5 and Table 6 suggest that providing feedback is the most powerful of the three activities explored within this analysis. (Note: this conclusion is based on a rough look at effect size within parallel simple regressions, not partial variation within a multiple regression.) But we did not determine *which* method of providing feedback delivers better results, nor the way that such an approach specifically improves the student experience. Is verbal group feedback more effective than written comments to individual students? Does the feedback improve the students' comprehension of the material, or the students' own sense of self-efficacy?

Although we demonstrated that the positive link between self-efficacy and student evaluation holds for both genders (with the exception of male professors and course design), we did not determine if male and female professors have the same definition of self-efficacy.

This paper assumes that a faculty member's self-report is an accurate reflection that can be compared to other people's self-reports. In other words, we assume that the lower average scores for female faculty members' self-efficacy demonstrate that they actually feel less confident than male faculty members. However, it is possible that different genders have different social norms and expectations for self-efficacy, undermining the ability to compare scores across genders (Ofem et al., 2019).

Finally, this study has a structural limitation. Although our data on faculty includes gender identity, we do not have any demographic data for the students who are making the evaluation. (The system to collect student evaluations in our sample intentionally separates the evaluation from student records to ensure that the evaluations are entirely and permanently anonymous.) We also lack more detailed demographic information on faculty, such as years of teaching experience or numbers of courses taught, which have helped to provide alternative explanations for performance. This gap in our data obviates several additional analyses that might have been useful to better understand the nuances of our finding in order to build stronger interventions that can improve self-efficacy to maximize student evaluations.

Despite these limitations, this paper concludes that pedagogical self-efficacy is an important theoretical and practical variable in the pursuit of higher student evaluations.

REFERENCES

Ambady, N. and Rosenthal, R. (1993) 'Predicting teacher evaluations from the thin slices of nonverbal behavior and physical attractiveness', *Journal of Personality and Social Psychology*, 64(3), pp. 431-441

Ashton, P. T. (1984) 'Teacher efficacy: A motivational paradigm for effective teacher education', *Journal of Teacher Education*, 35(5), pp. 28-32.

Apperson, J. M., Law, E.L., and Scepansky, J.A. (2006) 'The impact of presentation graphics on students' experience in the classroom', *Computers & Education*, 47(1), pp. 116-126.

Bandura, A. (1977) 'Self-efficacy: toward a unifying theory of behavioral change', *Psychological Review*, 84(2), pp. 191-215.

Bandura, A. (2001) 'Social cognitive theory: An agentic perspective', *Annual Review of Psychology*, 52, pp. 1–26.

Bandura, A. (2006) *Guide for Constructing Self-Efficacy Scales*. In F. Pajares and T. Urdan (Eds.). Greenwich, CT: Information Age Publishing.

Bandura, A. (2008) 'An agentic perspective on positive psychology', *Positive psychology, 1*, pp. 167-196.

BarNir, A., Watson, W. E., and Hutchins, H. M. (2011) 'Mediation and moderated mediation in the relationship among role models, self-efficacy, entrepreneurial career intention, and gender', *Journal of Applied Social Psychology*, 41(2), pp. 270-297.

Basow, S. A. (1995) 'Student evaluations of college professors: When gender matters', *Journal of Educational Psychology, 87*(4), pp. 656–665.

Beleche, T., Fairris, D., and Marks, M. (2012) 'Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test', *Economics of Education Review*, 31(5), pp. 709-719.

Bleicher, R. E. (2004) 'Revisiting the STEBI-B: Measuring self-efficacy in preservice elementary teachers', *School Science and Mathematics*, 104(8), pp. 383-391.

Boring, A., Ottoboni, K., and Stark, P. B. (2016) Student evaluations of teaching (mostly) do not measure teaching effectiveness, *Science Open Research*.

Centra, J.A. (1977) 'Faculty Development Practices', *New Directions for Higher Education*, 17(1977), pp. 49-55.

Centra, J.A. (1993) *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness.* San Francisco: Jossey-Bass.

Centra, J. A., and Gaubatz, N. B. (2000) 'Is there gender bias in student evaluations of teaching?', *Journal of Higher Education*, 71(1), pp.17–33.

Centra, J.A. (2003) 'Will Teachers Receive Higher Student Evaluations by Giving Higher Grades and Less Course Work?', *Research in Higher Education*, 44(5), pp. 495–518

Chen, C. C., Greene, P. G., and Crick, A. (1998) 'Does entrepreneurial self-efficacy distinguish entrepreneurs from managers?', *Journal of Business Venturing*, 13(4), pp. 295-316.

Charmaz, K. (2014) *Constructing Grounded Theory* (2nd ed.). Thousand Oaks, CA: Sage.

Cohen, P.A. (1980) 'Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings', *Research in Higher Education*, 13(4), pp. 321-341.

Cohen, P.A. (1981) 'Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies', *Review of Educational Research*, 51(3), pp. 281–309.

Collingridge, D. S., and Gantt, E. E. (2008) 'The quality of qualitative research', *American Journal of Medical Quality*, 23(5), pp. 389–395.

Corry, M., and Stella, J. (2018) 'Teacher self-efficacy in online education: a review of the literature', *Research in Learning Technology*, 26.

Cronbach, L. J., and Meehl, P. E. (1955) 'Construct validity in psychological tests', *Psychological Bulletin*, 52(4), pp. 281-302.

De Noble, A. F., Jung, D., and Ehrlich, S. B. (1999) 'Entrepreneurial self-efficacy: The development of a measure and its relationship to entrepreneurial action', *Frontiers of Entrepreneurship Research*, 1999, pp. 73-87.

Dellinger, A.B., Bobbett, J.J., Olivier, D.F., and Ellett, C.D. (2008) 'Measuring teachers' self-efficacy beliefs: Development and use of the TEBS-Self', *Teaching and Teacher Education*, 24(3), pp. 751-766.

Dickey, D., and Pearson, C. (2005) 'Recency Effect in College Student Course Evaluations', *Practical Assessment Research & Evaluation*, 10(6), pp. 1-10.

Evans, C., and Waring, M. (2011) 'Student teacher assessment feedback preferences: The influence of cognitive styles and gender', *Learning and Individual Differences*, 21(3), pp. 271-280.

Feldman, K. (1989) 'The Association between Student Ratings of Specific Instructional Dimensions and Student Achievement: Refining and Extending the Synthesis of Data from Multisection Validity Studies', *Research in Higher Education*, 30(6), pp. 583-645.

Field, A. (2018) *Discovering statistics using IBM SPSS Statistics*. 5th edn. London: Sage

Gibson, S., and Dembo, M. H. (1984) 'Teacher efficacy: A construct validation', *Journal of Educational Psychology*, 76(4), pp. 569-582.

Griffin, B.W. (2004) 'Grading leniency, grade discrepancy, and student ratings of instruction', *Contemporary Educational Psychology*, 29(4), pp. 410–25.

Gump, S.E. (2007) 'Student evaluations of teaching effectiveness and the leniency hypothesis: A literature review', *Educational Research Quarterly*, 30(3), pp. 56–6.

Guskey, T R. (1988) 'Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation', *Teaching and Teacher Education*, 4(1), pp. 63-69.

Hackman, M.Z., and Walker, K.B. (2009) 'Instructional communication in the televised classroom: The effects of system design and teacher immediacy on student learning and satisfaction', *Communication Education*, 39(3), pp. 196-206.

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2013) *Multivariate Data Analysis*: Global Edition, Seventh Edition: Pearson Education.

Hamermesh, D. S., and A. Parker. (2005) 'Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity', *Economics of Education Review*, 24(4), pp. 369-376.

Hmieleski, K. M., and Baron, R. A. (2008) 'When does entrepreneurial self-efficacy enhance versus reduce firm performance?', *Strategic Entrepreneurship Journal*, 2(1), pp. 57-72.

Hornstein, H. A., and Law, H. F. E. (2017) 'Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance', *Cogent Education,* 4(1).

Isaac C., Lee, B., & Carnes, M. (2009). Interventions that affect gender bias in hiring: A systematic review. *Academic Medicine, 84*(10), 1440-1446.

Jenkins, D.M. (2013) 'Exploring Instructional Strategies in Student Leadership Development Programming', *Journal of Leadership Studies,* 6(4), pp. 48-62.

Koon, J., and Murray, H. G. (1995) 'Using multiple outcomes to validate student ratings of overall teacher effectiveness', *Journal of Higher Education,* 66(1), pp. 61-81.

Ladd, T., Hind, P., and Lawrence, J. (2018) 'Entrepreneurial orientation, Waynesian self-efficacy for searching and marshaling, and intention across gender and region of origin', *Journal of Small Business & Entrepreneurship,* pp. 1-21.

Linse, A. R. (2017) 'Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees', *Studies in Educational Evaluation,* 54, pp. 94-106.

MacNell, L., Driscoll, A., and Hunt, A.N. (2015) 'What's in a Name: Exposing Gender Bias in Student Ratings of Teaching', *Innovative Higher Education,* 40(4), pp. 291-303.

Marsh, H. W. (1987) 'Students' evaluation of university teaching: Research findings, methodological issues, and directions for future research', *International Journal of Educational Research,* 11(3), pp. 253-388.

Marsh H.W. (2007) Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases and Usefulness. In: Perry R.P., Smart J.C. (eds) *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. Dordrecht: Springer.

Martin, A.E and Phillips, K.W. (2017) 'What "blindness" to gender differences helps women see and do: Implications for confidence, agency, and action in male-dominated environment', *Organizational Behaviour and Human Decision Processes,* 142, pp. 28-44

McGee, J. E., Peterson, M., Mueller, S. L., and Sequeira, J. M. (2009) 'Entrepreneurial self-efficacy: refining the measure', *Entrepreneurship Theory and Practice,* 33(4), pp. 965-988.

McGee, J. E., and Peterson, M. (2017) 'The Long-Term Impact of Entrepreneurial Self-Efficacy and Entrepreneurial Orientation on Venture Performance', *Journal of Small Business Management,* 57(3), pp. 720-737.

McPherson, M. A. (2006) 'Determinants of How Students Evaluate Teachers', *The Journal of Economic Education,* 37(1), pp. 3-20.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences, 109*(41), 16474-16479.

Ofem, B., Polizzi, S. J., Rushton, G. T., Beeth, M., Couch, B., Roehrig, G., and Sheppard, K. (2019) 'Gender Effects on Perceived Professional Mastery: Evidence from STEM Teachers', Paper presented at the Academy of Management Proceedings.

Pajares, M.F. (1992) 'Teachers' Belief and Educational Research: Cleaning up a messy construct', *Review of Educational Research,* 62(3), pp. 307-332.

Pillai, R., and Williams, E. (2004) 'Transformational leadership, self-efficacy, group cohesiveness, commitment, and performance', *Journal of Organizational Change Management,* 17(2), pp. 144-159.

Riggs, I. M., and Enochs, L. G. (1990) 'Toward the development of an elementary teacher's science teaching efficacy belief instrument', *Science Education,* 74(6), pp. 625-637.

Rivera, L. A., and Tilcsik, A. (2019) 'Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation', *American Sociological Review,* 84(2), pp. 248-274.

Saldana, J. (2011) *The Coding Manual for Qualitative Researchers*. London: Sage.

Spooren, P., Brock, B., and Mortelmans, D. (2013) 'On the Validity of Student Evaluation of Teaching', *Review of Educational Research,* 83(4), pp. 598-642.

Spooren, P., and Mortelmans, D. (2007) 'Teacher professionalism and student evaluation of teaching: will better teachers receive higher ratings and will better students give higher ratings?', *Educational Studies,* 32(2), pp. 201-214.

Staples, D. S., Hulland, J.S., and Higgins, C.A. (1999) 'A Self-Efficacy Theory Explanation for the Management of Remote Workers in Virtual Organizations', *Organization Science,* 10(6), pp. 758-776.

Stehle, S., Spinath, B., and Kadmon, M. (2012) 'Measuring Teaching Effectiveness: Correspondence Between Students' Evaluations of Teaching and Different Measures of Student Learning', *Research in Higher Education,* 53(8), pp. 888-904.

Steiner., D. D. and Rain, J.S. (1989) 'Immediate and delayed primacy and recency effects in performance evaluation', *Journal of Applied Psychology,* 74(1), pp. 136-142.

Stevens, J. P. (1992) *Applied Multivariate Statistics for the Social Sciences*. Hillsdale, NJ: Erbaum

Syed, M., and Nelson, S. C. (2015) 'Guidelines for establishing reliability when coding narrative data', *Emerging Adulthood,* 3(6), pp. 375–387.

Terrell J., Kofink, A., Middleton, J., Rainear, C., Murphy-Hill, E., Parnin, C., & Stallings, J. (2017). Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science, 3*.

Touchton, M. (2015) 'Flipping the Classroom and Student Performance in Advanced Statistics: Evidence from a Quasi-Experiment', *Journal of Political Science Education,* 11(1), pp. 28-44.

Tschannen-Moran, M., and Hoy, A.W. (2001) 'Teacher efficacy: Capturing an elusive construct', *Teaching and Teacher Education*, 17(7), pp. 783-805.

Urquhart, C. (2001) *An encounter with grounded theory: Tackling the practical and philosophical issue*s. In E. Trauth (Ed.), Qualitative research in IS: Issues and trends (pp. 104–140). Hershey, PA: Idea Group Publishing.

Urquhart, C., Lehmann, H., and Myers, M. D. (2010) 'Putting the 'theory' back into grounded theory: Guidelines for grounded theory studies in information systems', *Information Systems Journal*, 20(4), pp. 357–381.

Uttl, B., White, C.A., and Gonzalez, D.W. (2017) 'Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related', *Studies in Educational Evaluation*, 54, pp. 22-42.

Wright, S. L., and Jenkins-Guarnieri, M.A. (2012) 'Student evaluations of teaching: combining the meta-analyses and demonstrating further evidence for effective use', *Assessment & Evaluation in Higher Education*, 37(6), pp. 683-699.

Young, S., Rush, L., and Shaw, D. (2009) 'Evaluating gender bias in ratings of university instructors' teaching effectiveness', *International Journal for the Scholarship of Teaching and Learning*, 3(2), pp.1-14.

Zhao, H., Seibert, S. E., and Hills, G. E. (2005) 'The mediating role of self-efficacy in the development of entrepreneurial intentions', *Journal of Applied Psychology*, 90(6), pp. 1265-1272.